

Supplementary Information 4: Comparison of our Method to Geneland

Perhaps the most widely used program to detect barriers to gene flow based on spatially explicit data is Geneland (Guillot *et al.* 2005). In a recent study that compares different methods to detect barriers, it was identified as one of the most potent methods (Safner *et al.* 2011).

Here we test Geneland on two kinds of simulated data sets. One of them fulfills the model assumptions of Geneland, i.e. a barrier but not further substructure of the populations on either side of the barrier. For the second scenario, we simulated a scenario of a barrier with additional isolation by distance.

Application Details

In all analysis, we ran geneland with an MCMC chain length of 10^5 . For analysis, we used a thinning of 100 and a burn-in of 200, and visually inspected summary statistics to ensure proper convergence. We usually used a fixed population number $K = 2$, and investigated whether Geneland can accurately cluster the two subpopulations on each side of the barrier.

Scenario I: Two Panmictic Populations

In this scenario, we simulated data-sets that met the model assumptions of Geneland. Two equally sized populations on both sides of the barrier were assumed to be panmictic units, i.e. for all individuals alleles were binomially drawn with means p_l and p_r . We simulated datasets of 400 diploid individuals with genotype information for 200 biallelic loci spaced on 20×20 grid. The overall mean allele frequencies for the individuals left and right were randomly drawn:

$$\begin{aligned} p_l &= 0.5 + \Delta p \\ p_r &= 0.5 + \Delta p + \Delta p_r, \end{aligned}$$

where Δp and Δp_r are random normal variables with standard deviation $\sigma = 0.1$.

We simulated 10 replicates. In all of them, Geneland was able to accurately infer the position of the barrier and assigned all individuals correctly (Fig. S1).

Scenario II: Barrier with Isolation by Distance

Second, we applied Geneland on data-sets that we have generated using our explicit spatial population genetics simulations. We simulated 10 replicates of a datasets of 400 individuals for each a complete ($\gamma = 0$) and a weak barrier ($\gamma = 0.1$) with moderate isolation by distance ($m = 0.006$, $Nbh = 4\pi 5 \approx 62.83$ and $\sigma(p) = 0.1$). As expected, this data shows a clear isolation by distance pattern (Fig. S2). In all 10 data-sets, Geneland fails to accurately estimate the barrier, but rather infers 2 patchily distributed subpopulations (Fig. S3). It also cannot infer the barrier if the population number K is not fixed, but estimated as well (Results not shown). In contrast, our method is able to infer the existence and also the strength of a barrier in these scenarios (Fig. S4).

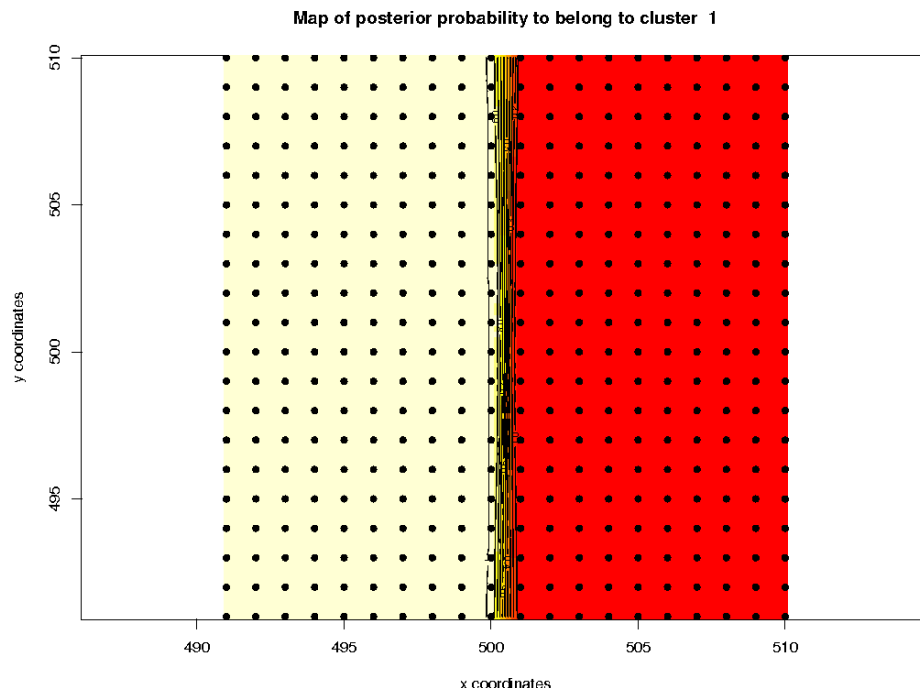


Figure S1: Geneland with no isolation by distance: This picture depicts a typical outcome when we applied Geneland to a model with a barrier but no further substructure (see main text). The figure visualizes posterior probability of population membership. Geneland is able to accurately infer the subdivision of the population into two subpopulations.

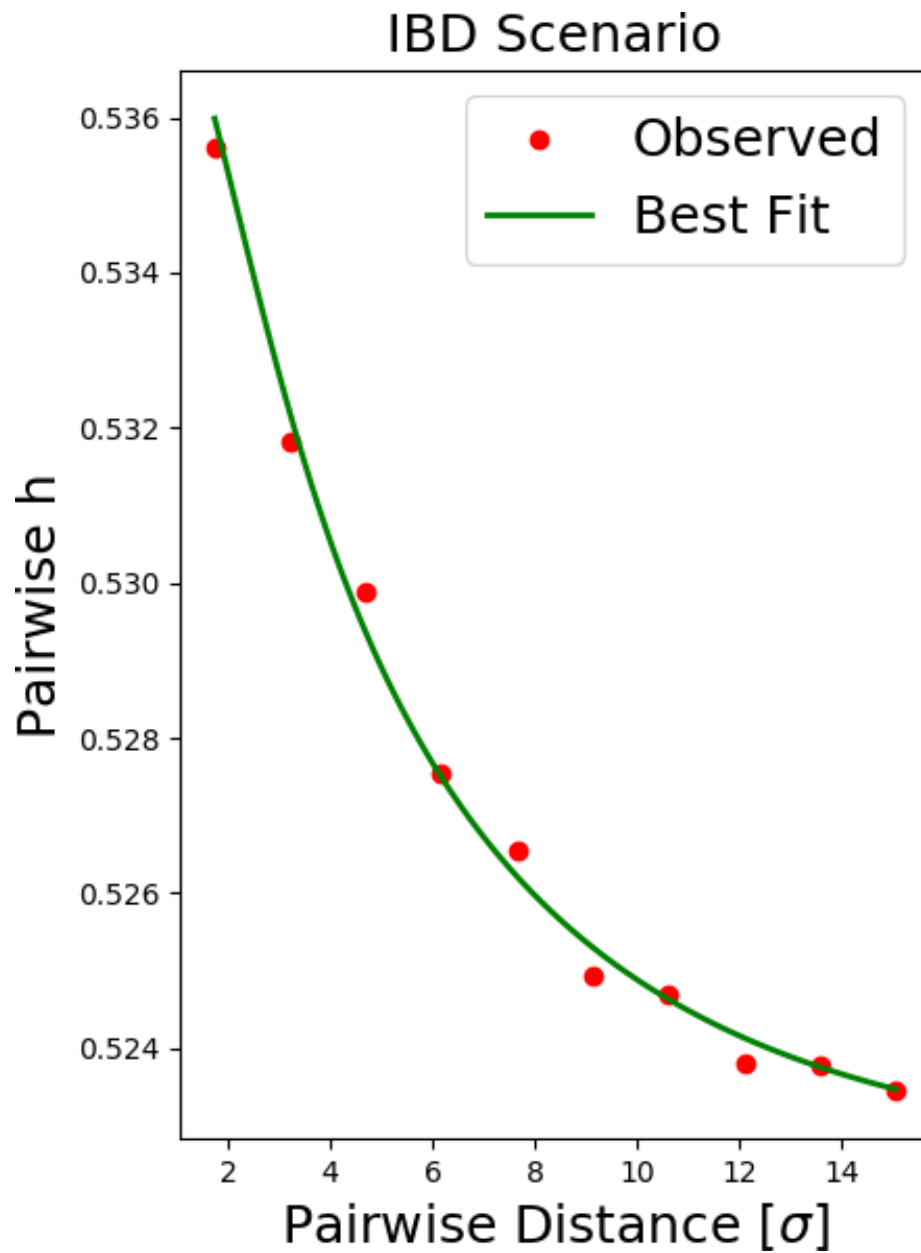


Figure S2: Isolation by Distance scenario: The decay of pairwise homozygosity in one dataset simulated under the Isolation by Distance scenarios.

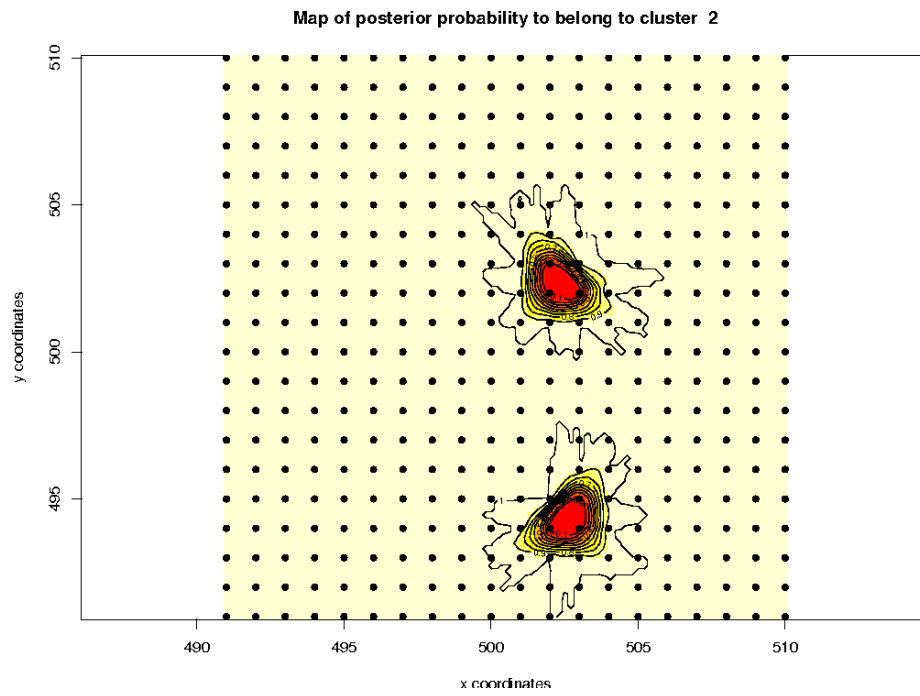


Figure S3: Geneland with isolation by distance: This picture shows a typical output of Geneland when applied to simulated data with a complete barrier and isolation by distance (see main text). The figure visualizes the posterior probability of population membership. Geneland fails to accurately infer the subdivision of the population into two subpopulations.

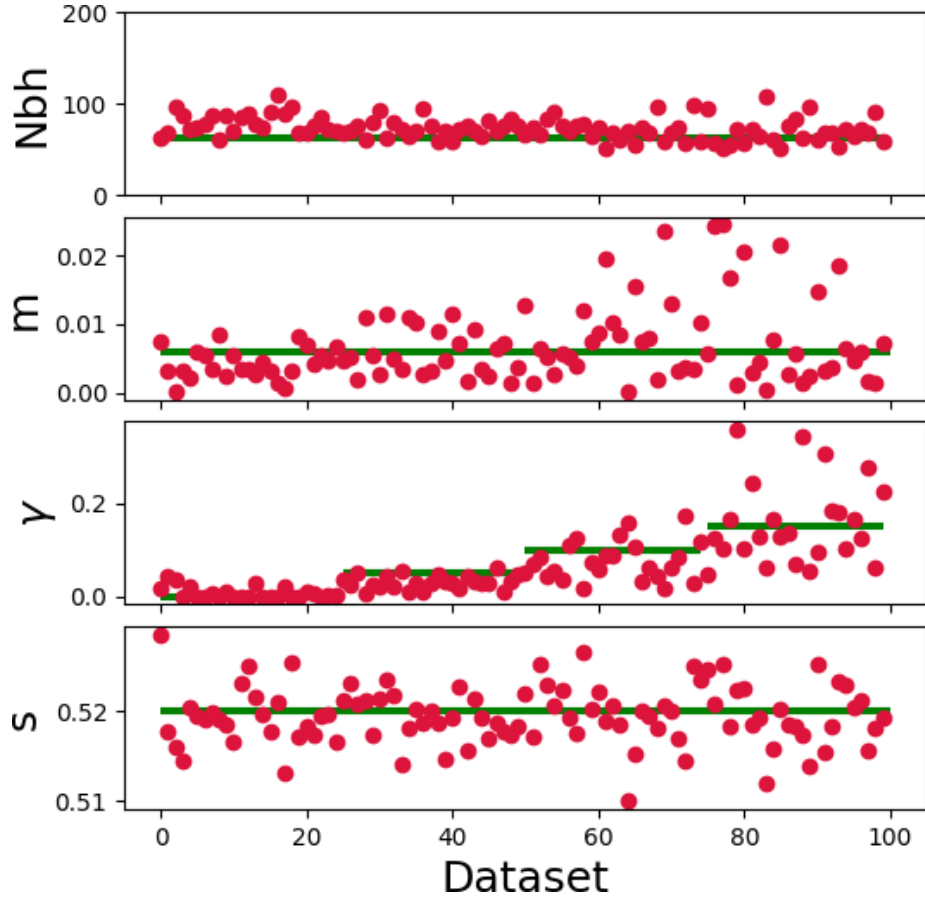


Figure S4: Our method on isolation by distance datasets: 25 replicates of four different barrier strengths were simulated $\gamma = 0, 0.05, 0.1, 0.15$. In all datasets Geneland failed to infer the barrier, whereas our method can estimate the strength of the barrier.

Conclusion

Our results show that Geneland is a powerful tool to infer a barrier when its model assumptions are met (i.e. population is structured into 2 subpopulations without further substructure). However, as observed previously (Safner *et al.* 2011), it fails in the scenario with additional isolation by distance, which we simulated under an explicit population genetics model. Our findings indicate that caution is warranted when applying Geneland to datasets with isolation by distance patterns. In particular, when the scale of isolation by distance observed on scales smaller than the geographic extension of the subpopulations, Geneland will have very limited power to detect a barrier. In contrast, our method works well in these cases. It can therefore be seen as a complementary approach to Geneland.

Literature Cited

- Guillot, G., A. Estoup, F. Mortier, and J. F. Cosson, 2005 A spatial statistical model for landscape genetics. *Genetics* **170**: 1261–1280.
- Safner, T., M. P. Miller, B. H. McRae, M.-J. Fortin, and S. Manel, 2011 Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics. *International Journal of Molecular Sciences* **12**: 865–889.